ISSN:2320-0790

An International Journal of Advanced Computer Technology

Development of content based image retrieval system using wavelet and Gabor transform

Manish Sharma[1], Vikash Shrivastava[2] and Vivek Shrivastava[3]

[1]Department of Information Technology, Student, ITM College Bhilwara Rajasthan India
[2]Senior Technical Architect, I-GATE Global Solution Limited, Noida U.P.,(India)
[3]Department of Information Technology, Head of Department, ITM College Bhilwara Rajasthan India

Abstract: A novel approach to image retrieval using color, texture and spatial information is proposed. The color information of an image is represented by the proposed color hologram, which takes into account both the occurrence of colors of pixels and the colors of their neighboring pixels. The proposed Fuzzy Color homogeneity, encoded by fuzzy sets, is incorporated in the color hologram computation. The texture information is described by the mean, variance and energy of wavelet decomposition coefficients in all sub bands. The spatial information is characterized by the class parameters obtained automatically from a unique unsupervised segmentation algorithm in combination with wavelet decomposition. Multi-stage filtering is applied to query processing to reduce the search range to speed up the query. Color homogram filter, wavelet texture filter, and spatial filter are used in sequence to eliminate images that are dissimilar to a query image in color, texture, and spatial information from the search ranges respectively. The proposed texture distance measure used in the wavelet texture filter considers the relationship between the coefficient value ranges and the decomposition levels, thus improving the retrieval performance.

Keywords: Content-based Image Retrieval (CBIR), Image Search, Image Descriptor, Wavelet Color-Spatial, Gabor Texture.

1. INTRODUCTION

As a result of the new communication technologies and the massive use of Internet in our society, the amount of audio-visual information available in digital format is increasing considerably. Therefore, it has been necessary to design some systems that allow us to describe the content of several types of multimedia information in order to search and classify them [1].

The audio-visual descriptors are in charge of the contents description. These descriptors have a good knowledge of the objects and events found in a video, image or audio and they allow the quick and efficient searches of the audio-visual content [4].

This system can be compared to the search engines for textual contents. Although it is certain, that it is relatively easy to find text with a computer, is much more difficult to find concrete audio and video parts. For instance, imagine somebody searching a scene of a happy person. The happiness is a feeling and it is not evident its shape, color and texture description in images [5].

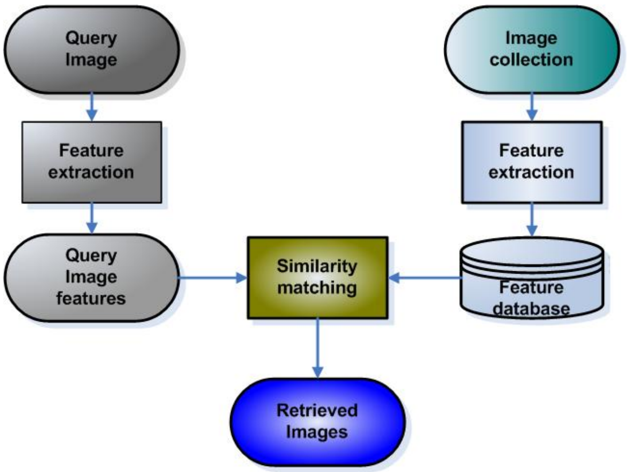
The description of the audio-visual content is not a superficial task and it is essential for the effective use of this type of archives. The standardization system that deals with audio-visual descriptors is the MPEG-7 (Motion Picture Expert Group - 7).

An improved method based on two layers image retrieval is proposed. The first layer retrieval extract edge of wavelet modulus maxima and use 7 moment invariants as the shape features and the averages and variances of the multi-scaling 2-D wavelet decomposition coefficients as the synthesis eigenvector [6]. The second one consists of the local color histograms. Difference weights are assigned to every block according to the importance of its spatial distribution. This method combines the advantages of the 7 moment invariants and the local spatial features, thereby achieving the better retrieval performance [7].

2.1 Content-Based Image Retrieval

When Information has been stored in textual format electronically, extracting required information automatically has never been hard. However when the information is held as an image the same is not at all true. Describing an Image in a concrete manner is a hard task. Although the Human can interpret what is seen, to be able to translate and teach the characteristics of the image to a machine is possible but it is a much difficult task. The retrieval of documents by a text based search engine does not require an understanding of the content. When retrieving images the same is not necessarily true. Most current algorithms fail to match high level concept of the image even if a similar image is provided to its low level concept. The problem here is relying more on visual similarities something which itself is highly subjective.

2 problems need to be resolved. Efficient and meaningful breaking of an image into parts based on features such as shape, color, texture & and large variance existing between low level features and high level expressions of an image. In CBIR, a description of the required content in terms of visual features of an image is required. Images can then be ranked with respect to their similarity to the description. These Images which most closely reflect the description are the ones to be retrieved.



"Content-based" means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web based image search engines rely purely on metadata and this produces a lot of garbage in the results. Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results.

3. Image retrieval

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.

3.1 Image Descriptor

This work has focused on techniques to learn SIFT like local image descriptors from large training datasets. Training data was obtained from existing 3D reconstructions, created using Photo Tourism (which is related to my earlier work on 3D reconstruction from unordered image collections), and multi-view stereo. Though these methods used local feature techniques to generate geometry, in principle any source (such as LIDAR) could be use to generate training data for learning.

Given a large dataset of corresponding image data, we seek to find a transformation (descriptor function) of that data that maximizes discrimination performance under a simple classifier (we use nearest neighbor). We have used two techniques to achieve this objective: 1) Powell minimization: we optimize parameterized descriptors to maximize ROC performance 2) LDA: we find a discriminate embedding that maximizes the ratio of between class / in class variance.

Results from these techniques are depicted above. The top figure shows optimal projections for LDE and an orthogonal variant (first two rows), the third row is PDA for comparison. The optimal linear discriminate features tend to focus on the centre of the image patch, and they tend to have the structure of circularly smoothed derivatives. The lower figure shows the pooling regions learnt using Powell minimization. These tend to have a foveated structure, a strategy found to be successful by other computer vision researchers (e.g. GLOH, DAISY).

The localized filter responses and foveated pooling regions also bear some resemblance to similar functionality in the human visual system.

3.2 Wavelet Transform

The most common form of transform image fusion is wavelet fusion. The wavelet transform decomposes the image into low-high, high-low, high-high spatial frequency bands at different scales and the low-low band at the coarsest scale. The low-low band contains the average image information whereas the other bands contain directional information due to spatial orientation [4]. Wavelet transform fusion is more formally defined by considering the wavelet transforms ω of the two input images $I_1(\chi, y)$ and $I_2(\chi, y)$ together with fusion rule Φ [12]. The inverse wavelet transform ω^{-1} is computed and the fused image $I(\chi, y)$ is reconstructed:

$$I(\chi, y) = \omega^{-1}(\Phi(\omega(I_1(\chi, y)), \omega(I_2(\chi, y))))$$

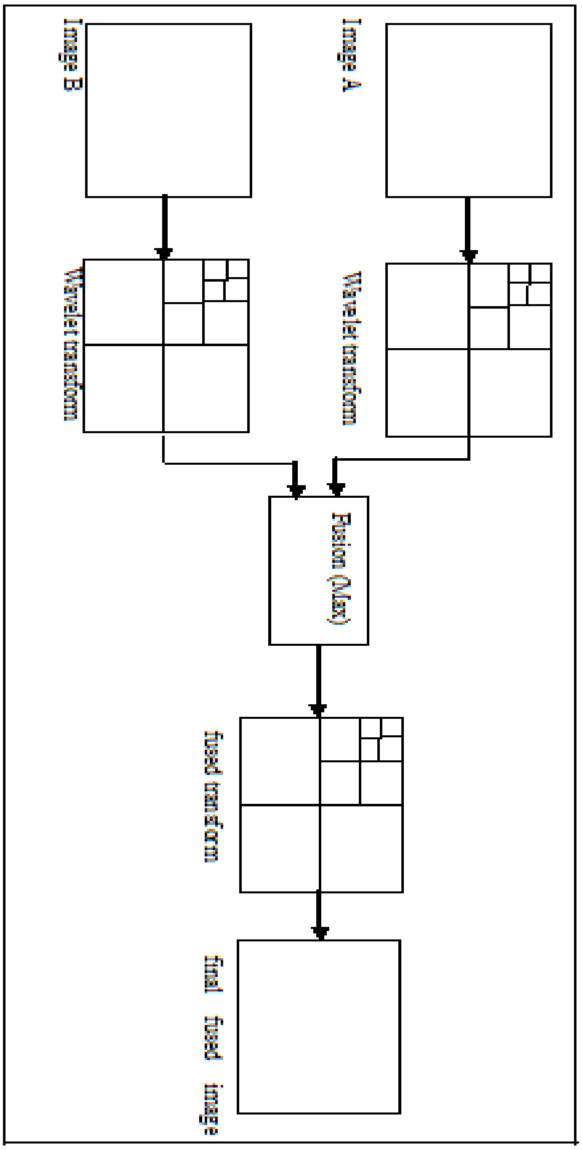
An alternative to fusion using pyramid based multi-resolution representations is fusion in the wavelet transform domain. The multiresolution wavelet representation is argued to be superior in several respects to that obtained with pyramidal methods:

1. Spatial orientation is introduced in the wavelet decomposition process, unlike pyramidal representations which do not include directional information.

2. The wavelet transform can be tailored to extract highly salient textures/edges while suppressing noise through the choice of the mother wavelet and high- and low-pass filters.

3. The different scales in the wavelet decomposition have a higher degree of independence than those in the pyramidal representations, which are correlated with each other.

The wavelet transform decomposes the image into low-high, high-low, and high-high spatial frequency bands at different scales and the low-low band at the coarsest scale. The L-L band contains the average image information whereas the other bands contain directional information due to spatial orientation. Higher absolute values of wavelet coefficients in the high bands correspond to salient features such as edges or lines. Since larger absolute transform coefficients correspond to sharper brightness changes, a good integration rule is to select, at every point in the transform domain, the coefficients whose absolute values are higher .



3.3 Gabor filter

In image processing, a Gabor filter, named after Dennis Gabor, is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The Gabor filters are self-similar: all filters can be generated from one mother wavelet by dilation and rotation.

3.4 Wavelet space

Gabor filters are directly related to Gabor wavelets, since they can be designed for a number of dilations and rotations. However, in general, expansion is not applied for Gabor wavelets, since this requires computation of bi-orthogonal wavelets, which may be very time-consuming. Therefore, usually, a filter bank consisting of Gabor filters with various scales and rotations is created. The filters are convolved with the signal, resulting in a so-called Gabor space. This process is closely related to processes in the primary visual cortex. Jones and Palmer showed that the real part of the complex Gabor function is a good fit to the receptive field weight functions found in simple cells in a cat's striate cortex.

The Gabor space is very useful in image processing applications such as optical character recognition, iris recognition and fingerprint recognition. Relations between activations for a specific spatial location are very distinctive between objects in an image. Furthermore, important activations can be extracted from the Gabor space in order to create a sparse object representation.

3.5 Gabor Texture

The characteristic of optimal joint resolution in both space and frequency suggests that these filters are appropriate operators for tasks requiring simultaneous measurement in these domains [2]. Texture discrimination is such a task. Computer application of a set of Gabor filters to a variety of textures found to be preattentively discriminable produces results in which differently textured regions are distinguished by first-order differences in the values measured by the filters. This ability to reduce the statistical complexity distinguishing differently textured region as well as the sensitivity of these filters to certain types of local features suggest that Gabor functions can act as detectors of certain "texton" types. The performance of the computer models suggests that cortical neurons with Gabor like receptive fields may be involved in preattentive texture discrimination [8].

4 Experiment Result

The most common evaluation measures used in IR (Information Retrieval) are precision and recall, usually presented as a precision-recall curve [6]. Precision denotes the ratio of retrieving an image that is relevant to the query, and recall indicates the ratio of the relevant images being retrieved, calculated as follows.

where *a* is the number of relevant images retrieved, *b* isthe number of irrelevant images retrieved, and *c* is the number of relevant images that were not retrieved.

Because precision and recall are not always the mostappropriate measures for evaluating IR, the precision and recall scores are often combined into a single measure of performance, known as the *F-score* [18]. Higher values of the F-score are obtained when both precision and recall are higher. The formula for calculating the F-score is:

The following experimental approach was adopted to evaluate the search results and quantify any improvement in the retrieval performance. Leave-one-out cross validation (LOO-CV) performance was applied to obtain more reliable estimates compared to previous experiments whose results were based on a small number of queries, for example MPEG-7 [9]. Thus, each image in the database was selected in turn as the query image, and queried against the remaining images.

5 CONCLUSION

This paper proposes a new efficient image descriptor that uses a combination of color features based on wavelet colorspatial information and texture features based on the Gabor wavelet texture of an image. In particular, when using a correlogram for the color features, more computational time is required than for a histogram-based approach. For this reason, as well as to better support a multi-resolution approach, we incorporated a wavelet transform, whose coefficients provide information that is independent of the original image resolution, and appropriately weight the LL, LH and HL sub-bands. Also, the use of color codebook helps to reduce the computing time needed. we proposed a new method for content based image retrieval. We tried to take into account the human visual system properties an incorporated the texture features to obtain better retrieval results. We utilized the dual-tree complex wavelet transform which is shown to be very efficient for extracting texture features from images and we showed its superiority over the Gabor filter. The experimental results show the efficiency of the proposed algorithm.

The results of experimental trials revealed that the proposed descriptor shows a significant improvement in retrieval effectiveness, especially in multi-resolution image searches. However, this approach does require additional computing time and storage space in memory buffer in comparison to other methods such as histogram-based approaches.

Ubiquitous computing environments provide a new and exciting area for information retrieval, but one of the challenges currently limiting many applications is how to best provide relevant information to users in real time. The key contribution of this paper lies in its use of a weighted combination of color and texture features to improve the performance of retrieval methods incorporating automated indexing for large image collections.

As for future work, there are two main avenues for further development the system to operate on smart phones such as the iPhone and Android and adding textual or semantically related information such as geo-location and user events to the existing algorithm to enable users to search for photographs associated with specific features in ubiquitous environments.

6 REFERENCE

[1] Mum-Kew Leong, Woo Chang, ISO/IEC PDTR 24800-1: JPSearch –Part 1: Framework and System Components, ISO/IEC JTC1/SC29/WG1N4203, Apr., 2007

[2] Arnol W.M. Smeulders, Marcel Warring, Simone Santini, Amarnarth Gupta, Ramesh Jain, Content-based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.12, 2000.

[3] B.S. Manjunath, Hens-Rainer Ohm, Vinod V. Vasudevan, Akio Yamada, Color and Texture Descriptors, IEEE Transactions on Circuits and System for Video Technology, vol.11, no.6, pp.703~715, June, 2001

[4] Ricardo Silva Torres, Alexandre Zavier Falcao, Content-based Image Retrieval: Theory and Applications, Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPO), Oct., 2006.

[5] Mun-Kew Leong, Wo Chang: Framework and System Components, ISO/IEC JTC1/SC29/WG1N3684, July, 2005.

[6] Vittorio Castelli, Lawrence D. Bergman: Image Databases – Search and Retrieval of Digital Imagery, Wiley Inter-Science, 2002

[7]. M.Swain and D. Ballard, “Color Indexing,” Int J.Comput. Vis., vol. 7, pp. 11-32, 1991.

[8]. A. W. M. Smeulders et al., “Content-Based Image Retrieval at the end of the Early Years,” IEEE Trans. On Pattern Anal. Machine Intell., vol. 22, pp.1349-1380, Dec. 2000.

[9]. M. Stricker and A. Dimai, “Spectral Covariance and Fuzzy Regions for Image Indexing,” Mach. Vis.Applicat., vol. 10, pp. 66-73, 1977

[10] James Z. Wang, "Integrated Region-Based Image Retrieval", Boston, Kluwer Academic Publishers, 2001

[11] J.R. Smith and S.F. Chang "Automated image retrieval using color and texture", Technical Report CU/CTR 408-95-14, Columbia University, July 1995.

[12] J.R. Smith and S.F. Chang. " Tools and techniques for color image retrieval", In Symposium on Electronic Imaging: Science and Technology - Storage & Retrieval for Image and Video Databases IV, volume 2670, San Jose, CA, February 1996. IS&T/SPIE.

[13] Sameer Antenna, Rangachar Caster, and Ramesh Jain. A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing and Retrieval of Images and Video. Pattern Recognition,35:945–965, 2002.

[14] I Daubechies, Ten Lectures on Wavelets, Capital City Press, 1992.

[15] Faloutsos, R. Barber, M. Flickner, J. Hafner,W. Niblack, D. Petkovic, W. Equitz, ``Efficient and effective querying by image content,' Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, vol. 3, no. 3-4, pp. 231-62, July 1994.